Predicting the Final League Tables of Domestic Football Leagues

Jan Van Haaren
Jesse Davis

Department of Computer Science, KU Leuven
jan.vanhaaren@cs.kuleuven.be

KU LEUVEN

DTAI
MACHINE LEARNING

The Guardian's predictions for the new Premier League season

Guardian writers predict the likely winners - and losers - of the new season

Vincent Kompany and his Manchester City team-mates lift last season's Premier League trophy. Can they repeat the feat this season? Photograph: Tom Jenkins

Source: The Guardian (retrieved 19 June 2015)

Phil McNulty

CHIEF FOOTBALL WRITER, BBC SPORT

More from Phil

Premier League 2014-15: Who will finish where?

13 August 2014
Last updated at 06:00 GMT

Share

Louis van Gaal has arrived at Old Trafford, Luis Suarez has left Anfield, and Manchester City and Chelsea have spent big again - so with the new Premier League season just days away from kick-off, here are my predictions for where each team will finish.

Source: BBC (retrieved 19 June 2015)

Motivation

- The prediction of **individual football matches**
  - has received significant attention due to the ever-growing interest in football betting
  - is a hard task due to matches being low-scoring
- The prediction of **final league tables**
  - has remained almost unexplored to date
  - is more interesting to managers and directors

Outline

- **Predictive ranking systems**
- Problem statement and approach
- Experimental evaluation
- Future work and conclusions

Predictive ranking systems

- Estimate **current strength** of teams based on their previous performances
- Provide a **complete ranking** of teams even when not all teams have played each other
- Have many possible **applications**:
  - seeding teams in tournament draws
  - granting work permits in the UK

Two types of rankings

- **Elo rankings** model absolute strength
  - *Traditional:* league-wide home-field advantage
  - *Advanced:* team-specific home-field advantage

- **Pi rankings** model relative strength

References: Elo (1978) and Constantinou and Fenton (2013)

Traditional Elo ranking

1. 

1575

2. 

1515

3. 

1490

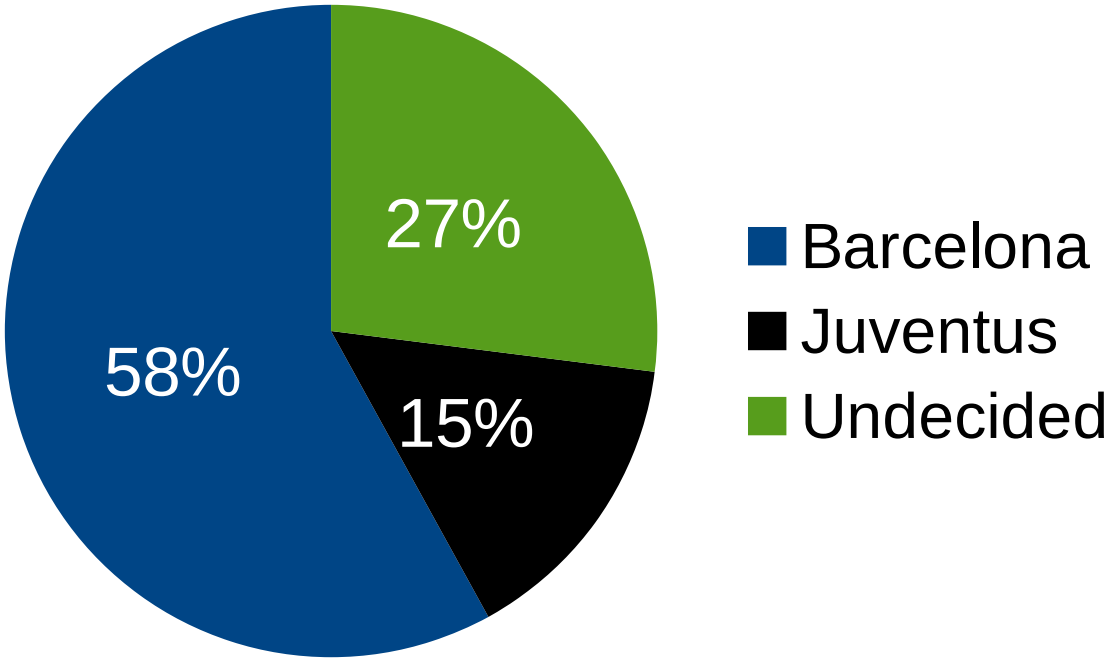
4. 

1410

Predicting the match outcome

1575
(+100)

1515



Updating the ranking

1575

(+100)

1515

3

15 points

1

Updating the ranking

1575

(+100)

1515

3 **15** points **1**

1 **35** points **3**

Updated Elo ranking

1.		1575	→	**1590**	(+15)
2.		1515	→	**1500**	(-15)
3.		1490	→	1490	
4.		1410	→	1410	

Advanced Elo ranking

1. 

1575

2. 

1515

3. 

1490

4. 

1410

Advanced Elo ranking

1. 

~~1575~~ → 1595 1555

2. 

~~1515~~ → 1535 1495

3. 

~~1490~~ → 1520 1460

4. 

~~1410~~ → 1405 1415

Pi ranking

1.		0.63	0.39
2.		0.33	0.23
3.		0.04	-0.35
4.		-0.48	-0.79

Predicting the match outcome

0.63

0.23

0.62 – 0.17 = 0.45

Predicted goal difference

Predicting the match outcome

0.63

0.23

0.62 – 0.17 = 0.45

Predicted goal difference

0.14 rating points

3 **1**

Updating the ranking

1. 

0.63 0.39

2. 

0.33 0.23

Updating the ranking

1. 

0.63

(+0.14)

0.77

0.39

2. 

0.33

0.23

(-0.14)

0.09

Updating the ranking

1. 

0.63 (+0.14) → 0.77

0.39 (+0.02) → 0.41

2. 

0.33 (-0.02) → 0.31

0.23 (-0.14) → 0.09

Updated Pi ranking

1.		0.77	0.41
2.		0.31	0.09
3.		0.04	-0.35
4.		-0.48	-0.79

Outline

- Predictive ranking systems
- **Problem statement and approach**
- Experimental evaluation
- Future work and conclusions

Problem statement

- **Given:** historical match results
  fixtures of upcoming matches

- **Predict:** final league table

Match result prediction

- **Elo rankings:** sample an outcome from the distribution over the possible outcomes

- **Pi ranking:** sample a result from a Poisson distribution over possible goal differences

League table prediction

1. Initialize the ranking
2. Predict the matches in chronological order
3. Update the ranking with predicted results
4. Compute league table based on predictions

League table prediction

- Match result prediction is **probabilistic**
- Hence, we perform a **Monte Carlo simulation**

Outline

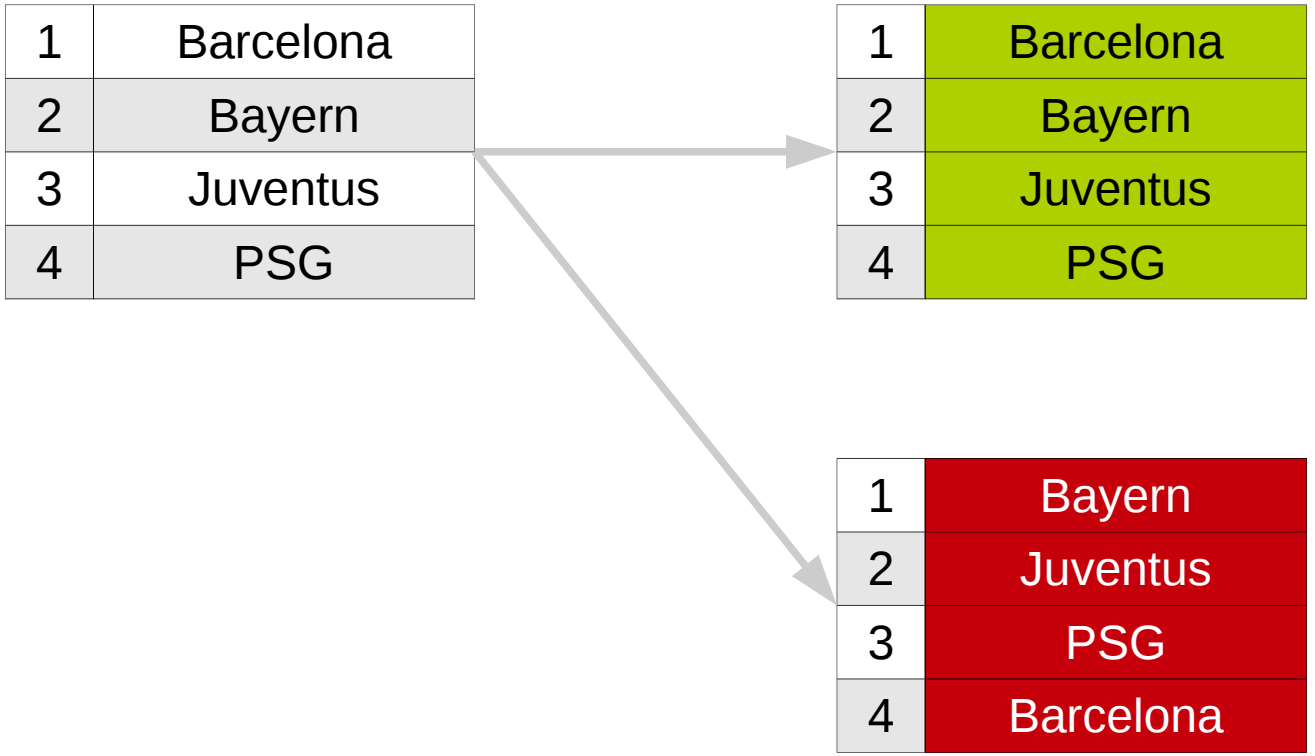
- Predictive ranking systems
- Problem statement
- **Experimental evaluation**
- Future work and conclusions

Experimental setup

- **Datasets:** (from football-data.co.uk)
  - Belgium, England, France, Germany, Italy, Spain, and The Netherlands
  - Seasons 2009/2010 through 2013/2014

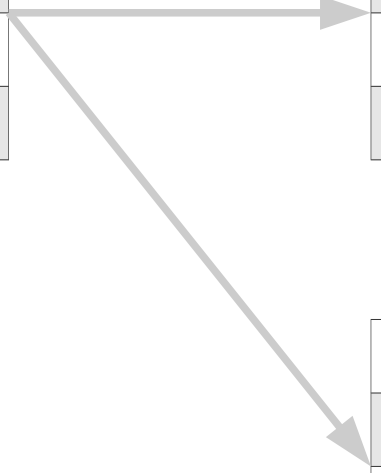
- **Evaluation metrics:**
  - Number of correctly predicted **absolute** positions
  - Number of correctly predicted **relative** positions
  - Rank correlation coefficients

Absolute accuracy



Relative accuracy

1	Barcelona
2	Bayern
3	Juventus
4	PSG



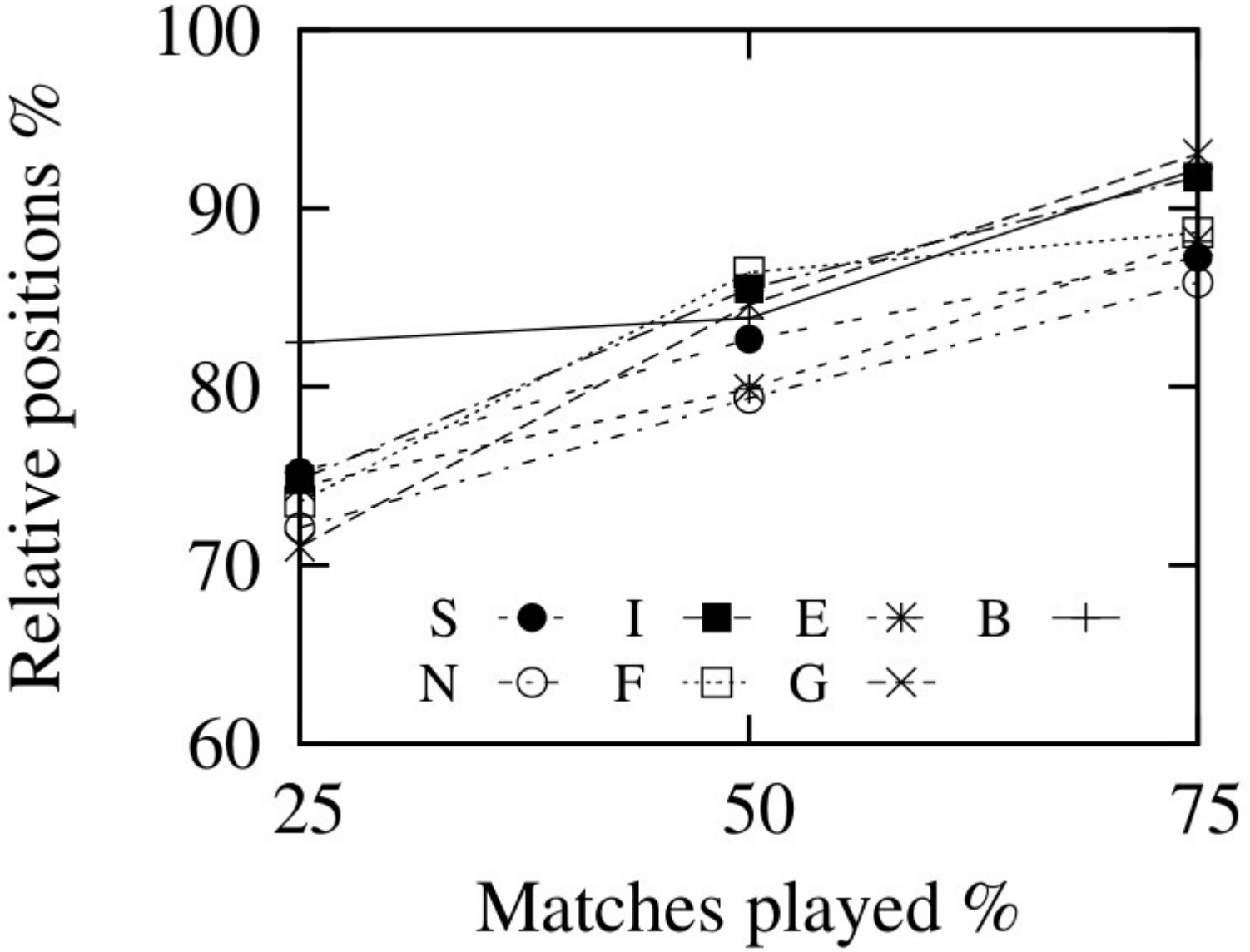
1	Barcelona
2	Bayern
3	Juventus
4	PSG

100%

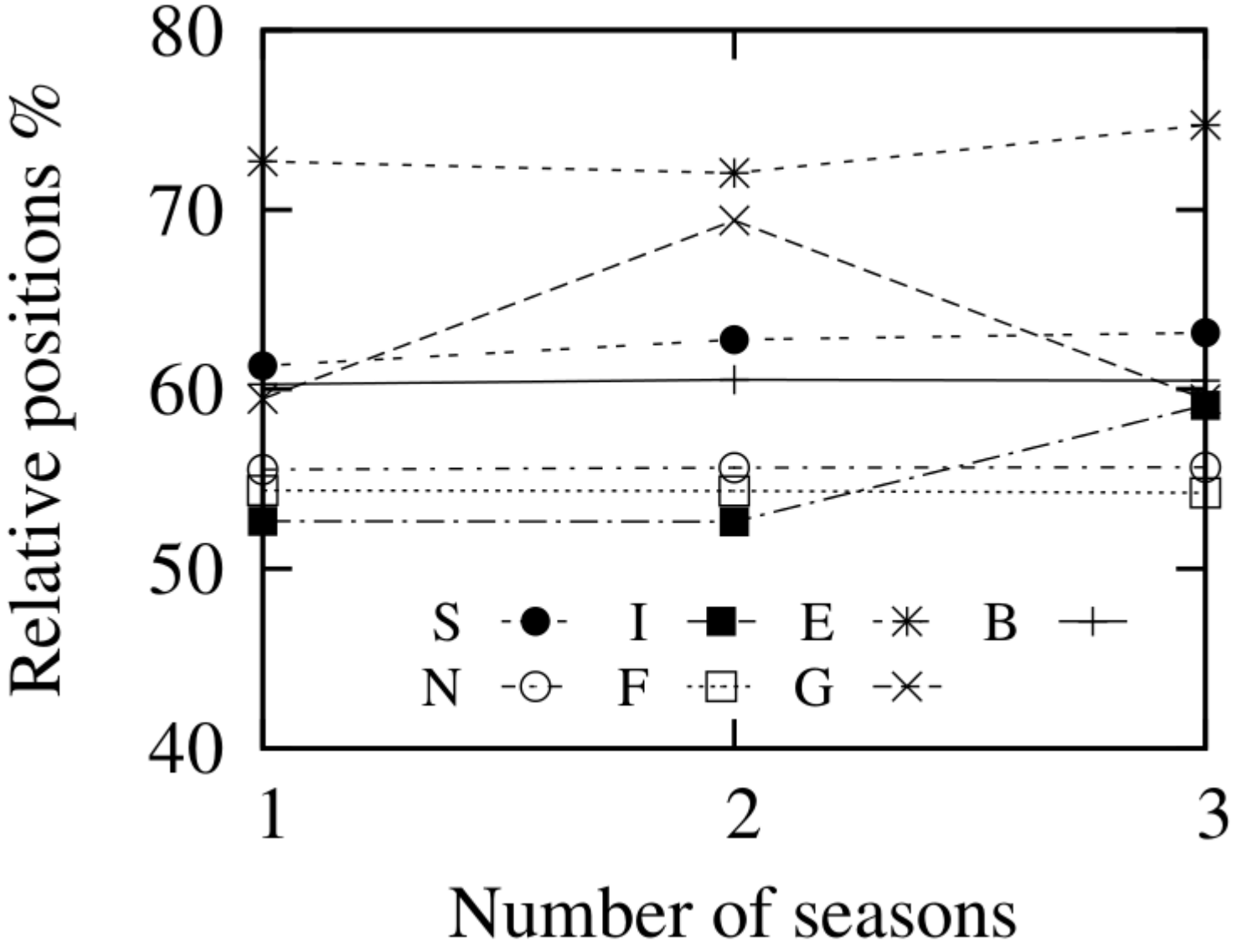
1	Bayern
2	Juventus
3	PSG
4	Barcelona

50%

Experimental evaluation



Experimental evaluation



Outline

- Predictive ranking systems
- Problem statement
- Experimental evaluation
- **Future work and conclusions**

Future work directions

- Devise more sophisticated **predictive rankings**
  - How to deal with the time aspect?
  - How to deal with missing players?
- Devise more meaningful **evaluation metrics**
  - Group positions into zones (e.g., CL or relegation)
  - Take importance of positions into account
- Devise **ensembles** of predictive rankings

Lessons learned

- The number of correctly predicted relative positions is a natural **evaluation metric**
- The **traditional Elo ranking** performs well in most of the settings that we considered
- The **Pi ranking** especially performs well when many historical match results are available

Predicting the Final League Tables of Domestic Football Leagues

Any questions, comments, suggestions, or ideas?

Jan Van Haaren – Jesse Davis
jan.vanhaaren@cs.kuleuven.be

KU LEUVEN

DTAI
MACHINE LEARNING